TransGAN: Two Pure Transformers Can Make One Strong GAN, and That Can Scale Up

Year: 2021 | Citations: 452 | Authors: Yifan Jiang, Shiyu Chang, Zhangyang Wang

---

Abstract

The recent explosive interest on transformers has suggested their potential to become powerful"universal"models for computer vision tasks, such as classification, detection, and segmentation. While those attempts mainly study the discriminative models, we explore transformers on some more notoriously difficult vision tasks, e.g., generative adversarial networks (GANs). Our goal is to conduct the first pilot study in building a GAN completely free of convolutions, using only pure transformer-based architectures. Our vanilla GAN architecture, dubbed TransGAN, consists of a memory-friendly transformer-based generator that progressively increases feature resolution, and correspondingly a multi-scale discriminator to capture simultaneously semantic contexts and low-level textures. On top of them, we introduce the new module of grid self-attention for alleviating the memory bottleneck further, in order to scale up TransGAN to high-resolution generation. We also develop a unique training recipe including a series of techniques that can mitigate the training instability issues of TransGAN, such as data augmentation, modified normalization, and relative position encoding. Our best architecture achieves highly competitive performance compared to current state-of-the-art GANs using convolutional backbones. Specifically, TransGAN sets new state-of-the-art inception score of 10.43 and FID of 18.28 on STL-10, outperforming StyleGAN-V2. When it comes to higher-resolution (e.g. 256 x 256) generation tasks, such as on CelebA-HQ and LSUN-Church, TransGAN continues to produce diverse visual examples with high fidelity and impressive texture details. In addition, we dive deep into the transformer-based generation models to understand how their behaviors differ from convolutional ones, by visualizing training dynamics. The code is available at https://github.com/VITA-Group/TransGAN.